Supporting Information

Using Street View Imagery to Predict Street-level Particulate Air Pollution

Meng Qi, Steve Hankey*

School of Public and International Affairs, Virginia Tech, 140 Otey Street, Blacksburg, Virginia 24061, United States

* Corresponding author, School of Public and International Affairs, Virginia Tech
Email: mengqi@vt.edu. Phone: 540.257.2332

The document contains additional information on the following topics:

Table of Contents

Description of data-driven feature selection

For the data-driven feature selection, we filtered predictor variables according to the statistical distribution of the variables. We used two data-filtering criteria on predictor variables in the "All variables" category. First, we set up thresholds for the mean of the predictor variables to filter out uncommon features for the overall study area. There are 1,350 potential predictor variables (150 classes × 9 buffer sizes) in total. We calculated the mean for each predictor variable at all aggregation locations as this metric could serve as a proxy for how common each feature is in the entire study area. According to the statistical distribution shown in Figure S1, we determined the thresholds of mean values to be 1×10^{-6}, 1×10^{-5} and 1×10^{-4}. The second filtering criteria was based on the nonzero percent of features among the aggregation locations to filter out uncommon features within measurement sites. Due to spatial discontinuity and heterogeneity, features along the monitoring routes varied as some features only appeared in a small number of places. For this criterion, the thresholds were set to be equally distributed: 0%, 20%, 40%, 60% and 80%. The aim of both criteria is to reduce uncommon features as these variables would harm the generalizability of our models. As a result, a total of 15 scenarios were created by combining these two filters for the data-driven feature selection.

Reduced buffer-feature GSV-only models

Since some grid centroids for air pollution prediction could not find GSV images nearby, we developed reduced buffer-feature models that avoid using central features or buffer features at low spatial radii. Specifically, we built models with the lowest buffer radius from 250m to 1,000m while the largest buffer size remained 2,000m. Table S7 lists the model parameters for the reduced buffer-feature GSV-only models as well as their model performance. All models used integrated feature selection with air pollution-related variables to prioritize model generalizability and interpretability. In addition, 81% of the grid centroids in the study area were able to find GSV images within 250m and the percentage was 94% when the radius extended to 500m. We explored two ways for applying the reduced buffer-feature models to make predictions for the 94% coverage of land area in Blacksburg. First, we developed an additive method which aimed to use models with the best possible spatial resolution for areas with limited GSV images. Specifically, for regions where GSV images were unavailable within 50 m but available within 250m, we used the 250m reduced buffer-feature models with buffer variables starting from 250m. For regions where GSV images were unavailable within 250 m but available within 500m, we used the 500m reduced buffer-feature models with buffer variables starting from 500m. Second, we used a simplified method, which simply used the reduced buffer-feature model with buffer variables starting from 500 m for all regions. Figure S8 shows the prediction maps using these methods as well as the estimates by the LUR models from our previous work[1] for comparison.

Results of spatial autocorrelation analysis and model uncertainty

Table S8 shows the results of Global Moran's I tests for our model residuals using different weights matrices and distance thresholds. The smallest radius (300m) is the minimum distance where each aggregation location finds at least one neighbor. In general, the model residuals demonstrate at least some spatial autocorrelation. This may be partly due to the fine spatial resolution of the mobile monitoring data and GSV images. To identify clusters of spatial autocorrelation, we performed a LISA (local indicator of spatial association) analysis. As shown in Figure S9, we chose the integrated buffer-feature models with air pollution-related variables for illustration. Generally, the high-high clusters (i.e., underprediction) are located where BC and PN concentrations are high while the low-low clusters (i.e., overprediction) are located where the observed concentrations are relatively low.

Figure S10 and S11 show model uncertainty by GSV-only models (integrated buffer-feature models with only air pollution-related variables for illustration) and previous LUR models[1]. Compared with the previous LUR models, GSV-only models show slightly higher standard error. The larger model uncertainty also tends to occur where BC and PN concentrations are at the higher or lower end of the observed distribution. As this study mainly aims to compare traditional GIS-derived predictors and our new GSV imagery-derived predictors, we chose to develop traditional step-wise linear regression models. A limitation of this choice is that our modeling framework does not account for spatial autocorrelation. A useful direction for future research is to use more advanced modeling approaches to address this issue, especially to improve model performance for extremely high or low concentration regions.

Table S1. Predictor variables in each category for the theory-driven feature selection

No.	Predictor variables	All variables	Theory-driven: Outdoor variables	Theory-driven: Air pollution-related variables	Reason to include/exclude variables from "Theory-driven: Air pollution related variables"
1	airplane	√	√		not found in GSV images in Blacksburg
2	animal	√	√		not found in GSV images in Blacksburg
3	apparel	√			indoor
4	arcade machine	√			indoor
5	armchair	√			indoor
6	ashcan	√			indoor
7	awning	√	√		no air pollutant emission
8	bag	√			indoor
9	ball	√			indoor
10	bannister	√			indoor
11	bar	√			indoor
12	barrel	√			indoor
13	base	√			indoor
14	basket	√			indoor
15	bathtub	√			indoor
16	bed	√			indoor
17	bench	√			indoor
18	bicycle	√	√	√	traffic
19	blanket	√			indoor
20	blind	√			indoor
21	boat	√	√		not found in GSV images in Blacksburg
22	book	√			indoor
23	bookcase	√			indoor
24	booth	√			indoor
25	bottle	√			indoor
26	box	√			indoor
27	bridge	√	√	√	traffic
28	buffet	√			indoor
29	building	√	√	√	pollutant dispersion
30	bulletin board	√			indoor
31	bus	√	√	√	traffic
32	cabinet	√			indoor
33	canopy	√	√		not found in GSV images in Blacksburg
34	car	√	√	√	traffic
35	case	√			indoor
36	ceiling	√			indoor
37	chair	√			indoor
38	chandelier	√			indoor
39	chest of drawers	√			indoor
40	clock	√			indoor
41	coffee table	√			indoor
42	column	√			indoor
43	computer	√			indoor
44	conveyer belt	√			indoor
45	counter	√			indoor
46	countertop	√			indoor
47	cradle	√			indoor
48	crt screen	√			indoor
49	curtain	√			indoor
50	cushion	√			indoor

No.	Predictor variables	All variables	Theory-driven: Outdoor variables	Theory-driven: Air pollution-related variables	Reason to include/exclude variables from "Theory-driven: Air pollution related variables"
51	desk	√			indoor
52	dirt track	√	√	√	dust
53	dishwasher	√			indoor
54	door	√			indoor
55	earth	√	√	√	dust/green space
56	escalator	√			indoor
57	fan	√			indoor
58	fence	√	√		no air pollutant emission
59	field	√	√	√	dust/green space
60	fireplace	√			indoor
61	flag	√	√		no air pollutant emission
62	floor	√			indoor
63	flower	√	√	√	green space
64	food	√			indoor
65	fountain	√	√		not found in GSV images in Blacksburg
66	glass	√			indoor
67	grandstand	√			indoor
68	grass	√	√	√	green space
69	hill	√	√	√	green space
70	hood	√			indoor
71	house	√	√	√	pollutant dispersion
72	hovel	√	√		not found in GSV images in Blacksburg
73	kitchen island	√			indoor
74	lake	√	√		not found in GSV images in Blacksburg
75	lamp	√			indoor
76	land	√	√	√	dust/green space
77	light	√			indoor
78	microwave	√			indoor
79	minibike	√	√	√	traffic
80	mirror	√			indoor
81	monitor	√			indoor
82	mountain	√	√	√	green space
83	ottoman	√			indoor
84	oven	√			indoor
85	painting	√			indoor
86	palm	√	√		not found in GSV images
87	path	√	√	√	traffic
88	person	√	√	√	possible emission source
89	pier	√	√		not found in GSV images in Blacksburg
90	pillow	√			indoor
91	plant	√	√	√	green space
92	plate	√			indoor
93	plaything	√			indoor
94	pole	√	√		no air pollutant emission
95	pool table	√			indoor
96	poster	√			indoor
97	pot	√			indoor
98	radiator	√			indoor
99	railing	√	√		no air pollutant emission
100	refrigerator	√			indoor
101	river	√	√		not found in GSV images in Blacksburg
102	road	√	√	√	traffic
103	rock	√	√	√	dust

No.	Predictor variables	All variables	Theory-driven: Outdoor variables	Theory-driven: Air pollution-related variables	Reason to include/exclude variables from "Theory-driven: Air pollution related variables"
104	rug	√			indoor
105	runway	√	√	√	traffic
106	sand	√	√	√	dust
107	sconce	√			indoor
108	screen	√			indoor
109	screen door	√			indoor
110	sculpture	√	√		no air pollutant emission
111	sea	√	√		not found in GSV images in Blacksburg
112	seat	√			indoor
113	shelf	√			indoor
114	ship	√	√		not found in GSV images in Blacksburg
115	shower	√			indoor
116	sidewalk	√	√	√	traffic
117	signboard	√	√	√	traffic
118	sink	√			indoor
119	sky	√	√	√	pollutant dispersion
120	skyscraper	√	√		not found in GSV images in Blacksburg
121	sofa	√			indoor
122	stage	√			indoor
123	stairs	√			indoor
124	stairway	√			indoor
125	step	√			indoor
126	stool	√			indoor
127	stove	√			indoor
128	streetlight	√	√	√	traffic
129	swimming pool	√	√		not found in GSV images in Blacksburg
130	swivel chair	√			indoor
131	table	√			indoor
132	tank	√			indoor
133	television receiver	√			indoor
134	tent	√	√		no air pollutant emission
135	toilet	√			indoor
136	towel	√			indoor
137	tower	√	√		no air pollutant emission
138	trade name	√	√		no air pollutant emission
139	traffic light	√	√	√	traffic
140	tray	√			indoor
141	tree	√	√	√	green space
142	truck	√	√	√	traffic
143	van	√	√	√	traffic
144	vase	√			indoor
145	wall	√	√	√	pollutant dispersion
146	wardrobe	√			indoor
147	washer	√			indoor
148	water	√	√		not found in GSV images in Blacksburg
149	waterfall	√	√		not found in GSV images in Blacksburg
150	windowpane	√	√		no air pollutant emission

Table S2. Descriptive statistics of features extracted from sampled GSV images within 2,000m of air quality monitoring locations

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
1	airplane	0.00E+00	0.00E+00	0.00E+00	3.87E-06	0.00E+00	4.49E-03	9.23E-05
2	animal	0.00E+00	0.00E+00	0.00E+00	3.01E-06	0.00E+00	4.93E-03	7.17E-05
3	apparel	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
4	arcade machine	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
5	armchair	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
6	ashcan	0.00E+00	0.00E+00	0.00E+00	1.57E-04	0.00E+00	1.90E-02	8.46E-04
7	awning	0.00E+00	0.00E+00	0.00E+00	2.74E-05	0.00E+00	2.60E-02	4.94E-04
8	bag	0.00E+00	0.00E+00	0.00E+00	3.23E-06	0.00E+00	3.84E-03	6.61E-05
9	ball	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
10	bannister	0.00E+00	0.00E+00	0.00E+00	1.42E-05	0.00E+00	1.32E-02	2.17E-04
11	bar	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
12	barrel	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
13	base	0.00E+00	0.00E+00	0.00E+00	2.30E-05	0.00E+00	1.63E-02	3.54E-04
14	basket	0.00E+00	0.00E+00	0.00E+00	7.22E-07	0.00E+00	2.13E-03	3.08E-05
15	bathtub	0.00E+00	0.00E+00	0.00E+00	1.00E-06	0.00E+00	8.10E-03	8.03E-05
16	bed	0.00E+00	0.00E+00	0.00E+00	2.98E-06	0.00E+00	1.71E-02	1.94E-04
17	bench	0.00E+00	0.00E+00	0.00E+00	4.50E-05	0.00E+00	1.46E-02	3.49E-04
18	bicycle	0.00E+00	0.00E+00	0.00E+00	6.68E-06	0.00E+00	6.95E-03	1.35E-04
19	blanket	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
20	blind	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
21	boat	0.00E+00	0.00E+00	0.00E+00	5.81E-06	0.00E+00	8.19E-03	1.33E-04
22	book	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
23	bookcase	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
24	booth	0.00E+00	0.00E+00	0.00E+00	6.39E-08	0.00E+00	7.00E-04	6.69E-06
25	bottle	0.00E+00	0.00E+00	0.00E+00	1.61E-07	0.00E+00	1.19E-03	1.21E-05
26	box	0.00E+00	0.00E+00	0.00E+00	9.68E-05	0.00E+00	2.39E-02	7.09E-04
27	bridge	0.00E+00	0.00E+00	0.00E+00	3.87E-04	0.00E+00	1.47E-01	3.74E-03
28	buffet	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
29	building	0.00E+00	1.02E-03	6.16E-03	1.93E-02	1.87E-02	5.60E-01	3.91E-02
30	bulletin board	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
31	bus	0.00E+00	0.00E+00	0.00E+00	1.22E-04	0.00E+00	1.45E-01	2.82E-03
32	cabinet	0.00E+00	0.00E+00	0.00E+00	2.48E-06	0.00E+00	7.83E-03	1.18E-04
33	canopy	0.00E+00	0.00E+00	0.00E+00	1.85E-06	0.00E+00	8.43E-03	9.41E-05
34	car	0.00E+00	1.54E-04	1.46E-03	7.82E-03	6.27E-03	1.84E-01	1.68E-02
35	case	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
36	ceiling	0.00E+00	0.00E+00	0.00E+00	3.75E-04	0.00E+00	3.93E-01	7.78E-03
37	chair	0.00E+00	0.00E+00	0.00E+00	6.11E-06	0.00E+00	7.26E-03	1.23E-04
38	chandelier	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
39	chest of drawers	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
40	clock	0.00E+00	0.00E+00	0.00E+00	2.47E-07	0.00E+00	7.59E-04	1.20E-05
41	coffee table	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
42	column	0.00E+00	0.00E+00	0.00E+00	3.53E-05	0.00E+00	4.49E-02	7.36E-04
43	computer	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
44	conveyer belt	0.00E+00	0.00E+00	0.00E+00	3.07E-07	0.00E+00	3.36E-03	3.21E-05
45	counter	0.00E+00	0.00E+00	0.00E+00	1.90E-06	0.00E+00	7.52E-03	9.73E-05
46	countertop	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
47	cradle	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
48	crt screen	0.00E+00	0.00E+00	0.00E+00	3.43E-08	0.00E+00	3.76E-04	3.59E-06
49	curtain	0.00E+00	0.00E+00	0.00E+00	4.89E-07	0.00E+00	4.95E-03	4.75E-05
50	cushion	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
51	desk	0.00E+00	0.00E+00	0.00E+00	2.31E-07	0.00E+00	1.34E-03	1.62E-05
52	dirt track	0.00E+00	0.00E+00	0.00E+00	1.10E-04	0.00E+00	6.09E-02	1.49E-03
53	dishwasher	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
54	door	0.00E+00	0.00E+00	0.00E+00	1.26E-05	0.00E+00	2.69E-02	3.60E-04
55	earth	0.00E+00	2.17E-04	5.88E-03	2.65E-02	3.24E-02	3.36E-01	4.52E-02
56	escalator	0.00E+00	0.00E+00	0.00E+00	7.37E-07	0.00E+00	3.90E-03	4.42E-05
57	fan	0.00E+00	0.00E+00	0.00E+00	6.17E-08	0.00E+00	6.64E-04	6.35E-06
58	fence	0.00E+00	0.00E+00	7.63E-05	3.76E-03	2.25E-03	1.98E-01	9.63E-03
59	field	0.00E+00	0.00E+00	0.00E+00	4.70E-03	3.30E-05	3.20E-01	1.99E-02
60	fireplace	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
61	flag	0.00E+00	0.00E+00	0.00E+00	9.14E-06	0.00E+00	2.03E-02	2.19E-04
62	floor	0.00E+00	0.00E+00	0.00E+00	1.46E-04	0.00E+00	1.34E-01	2.50E-03
63	flower	0.00E+00	0.00E+00	0.00E+00	3.90E-05	0.00E+00	2.47E-02	5.15E-04
64	food	0.00E+00	0.00E+00	0.00E+00	1.85E-06	0.00E+00	1.43E-02	1.41E-04
65	fountain	0.00E+00	0.00E+00	0.00E+00	2.20E-06	0.00E+00	5.87E-03	8.58E-05
66	glass	0.00E+00	0.00E+00	0.00E+00	1.84E-09	0.00E+00	2.01E-05	1.92E-07
67	grandstand	0.00E+00	0.00E+00	0.00E+00	8.38E-06	0.00E+00	3.83E-02	4.29E-04
68	grass	0.00E+00	5.63E-02	1.05E-01	1.08E-01	1.53E-01	3.51E-01	6.73E-02
69	hill	0.00E+00	0.00E+00	0.00E+00	2.27E-04	0.00E+00	6.62E-02	2.22E-03
70	hood	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
71	house	0.00E+00	0.00E+00	2.54E-04	7.09E-03	8.95E-03	1.21E-01	1.36E-02
72	hovel	0.00E+00	0.00E+00	0.00E+00	9.78E-06	0.00E+00	2.36E-02	3.69E-04
73	kitchen island	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
74	lake	0.00E+00	0.00E+00	0.00E+00	3.91E-06	0.00E+00	1.59E-02	1.97E-04
75	lamp	0.00E+00	0.00E+00	0.00E+00	2.20E-08	0.00E+00	1.02E-04	1.31E-06
76	land	0.00E+00	0.00E+00	0.00E+00	3.53E-06	0.00E+00	1.24E-02	1.27E-04
77	light	0.00E+00	0.00E+00	0.00E+00	3.18E-07	0.00E+00	1.15E-03	1.72E-05
78	microwave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
79	minibike	0.00E+00	0.00E+00	0.00E+00	9.24E-06	0.00E+00	1.52E-02	2.48E-04
80	mirror	0.00E+00	0.00E+00	0.00E+00	9.31E-08	0.00E+00	1.02E-03	9.74E-06
81	monitor	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
82	mountain	0.00E+00	0.00E+00	0.00E+00	5.64E-04	2.32E-05	1.25E-01	4.07E-03
83	ottoman	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
84	oven	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
85	painting	0.00E+00	0.00E+00	0.00E+00	5.47E-07	0.00E+00	2.56E-03	2.96E-05
86	palm	0.00E+00	0.00E+00	0.00E+00	9.12E-06	0.00E+00	2.37E-02	3.35E-04
87	path	0.00E+00	0.00E+00	2.44E-06	4.81E-03	2.80E-03	2.13E-01	1.26E-02
88	person	0.00E+00	0.00E+00	0.00E+00	1.17E-04	3.05E-06	3.71E-02	7.35E-04
89	pier	0.00E+00	0.00E+00	0.00E+00	2.23E-08	0.00E+00	2.44E-04	2.33E-06
90	pillow	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
91	plant	0.00E+00	4.80E-04	4.87E-03	1.41E-02	1.72E-02	2.47E-01	2.31E-02
92	plate	0.00E+00	0.00E+00	0.00E+00	1.16E-07	0.00E+00	6.68E-04	7.14E-06
93	plaything	0.00E+00	0.00E+00	0.00E+00	3.14E-07	0.00E+00	2.18E-03	2.34E-05
94	pole	0.00E+00	0.00E+00	0.00E+00	3.07E-04	5.49E-05	2.80E-02	1.12E-03
95	pool table	0.00E+00	0.00E+00	0.00E+00	6.91E-09	0.00E+00	7.02E-05	6.73E-07
96	poster	0.00E+00	0.00E+00	0.00E+00	4.89E-07	0.00E+00	2.31E-03	2.72E-05
97	pot	0.00E+00	0.00E+00	0.00E+00	8.22E-06	0.00E+00	3.76E-03	8.72E-05
98	radiator	0.00E+00	0.00E+00	0.00E+00	3.27E-07	0.00E+00	3.58E-03	3.42E-05
99	railing	0.00E+00	0.00E+00	0.00E+00	7.83E-05	0.00E+00	4.96E-02	1.01E-03
100	refrigerator	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
101	river	0.00E+00	0.00E+00	0.00E+00	1.22E-04	0.00E+00	5.32E-02	1.50E-03
102	road	0.00E+00	2.08E-01	2.63E-01	2.58E-01	3.09E-01	4.61E-01	7.29E-02
103	rock	0.00E+00	0.00E+00	0.00E+00	2.40E-04	0.00E+00	3.53E-02	1.45E-03
104	rug	0.00E+00	0.00E+00	0.00E+00	5.91E-07	0.00E+00	4.30E-03	4.39E-05
105	runway	0.00E+00	0.00E+00	0.00E+00	4.60E-06	0.00E+00	9.24E-03	1.65E-04
106	sand	0.00E+00	0.00E+00	0.00E+00	8.68E-05	0.00E+00	5.84E-02	1.24E-03
107	sconce	0.00E+00	0.00E+00	0.00E+00	1.18E-07	0.00E+00	8.26E-04	8.52E-06
108	screen	0.00E+00	0.00E+00	0.00E+00	1.68E-07	0.00E+00	1.40E-03	1.40E-05
109	screen door	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
110	sculpture	0.00E+00	0.00E+00	0.00E+00	1.97E-06	0.00E+00	4.96E-03	6.49E-05
111	sea	0.00E+00	0.00E+00	0.00E+00	1.07E-06	0.00E+00	3.86E-03	4.27E-05
112	seat	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
113	shelf	0.00E+00	0.00E+00	0.00E+00	9.48E-08	0.00E+00	1.04E-03	9.92E-06
114	ship	0.00E+00	0.00E+00	0.00E+00	1.51E-06	0.00E+00	1.64E-02	1.56E-04
115	shower	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
116	sidewalk	0.00E+00	0.00E+00	8.39E-04	1.32E-02	1.53E-02	3.56E-01	2.53E-02
117	signboard	0.00E+00	1.47E-05	1.92E-04	6.30E-04	5.93E-04	4.69E-02	1.62E-03
118	sink	0.00E+00	0.00E+00	0.00E+00	4.34E-08	0.00E+00	3.03E-04	3.33E-06
119	sky	7.66E-03	1.98E-01	2.95E-01	2.81E-01	3.76E-01	4.85E-01	1.15E-01
120	skyscraper	0.00E+00	0.00E+00	0.00E+00	6.30E-09	0.00E+00	6.90E-05	6.59E-07
121	sofa	0.00E+00	0.00E+00	0.00E+00	3.24E-08	0.00E+00	3.55E-04	3.39E-06
122	stage	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
123	stairs	0.00E+00	0.00E+00	0.00E+00	9.92E-05	0.00E+00	1.98E-02	6.67E-04
124	stairway	0.00E+00	0.00E+00	0.00E+00	1.58E-05	0.00E+00	1.45E-02	2.40E-04
125	step	0.00E+00	0.00E+00	0.00E+00	2.87E-07	0.00E+00	2.14E-03	2.21E-05
126	stool	0.00E+00	0.00E+00	0.00E+00	2.29E-09	0.00E+00	2.44E-05	2.33E-07
127	stove	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
128	streetlight	0.00E+00	0.00E+00	0.00E+00	1.84E-04	1.08E-04	8.89E-03	4.77E-04
129	swimming pool	0.00E+00	0.00E+00	0.00E+00	1.20E-06	0.00E+00	1.23E-02	1.17E-04
130	swivel chair	0.00E+00	0.00E+00	0.00E+00	1.90E-09	0.00E+00	2.08E-05	1.98E-07
131	table	0.00E+00	0.00E+00	0.00E+00	5.41E-06	0.00E+00	9.82E-03	1.46E-04
132	tank	0.00E+00	0.00E+00	0.00E+00	9.96E-07	0.00E+00	5.30E-03	5.65E-05
133	television receiver	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
134	tent	0.00E+00	0.00E+00	0.00E+00	2.69E-05	0.00E+00	1.37E-01	1.34E-03
135	toilet	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
136	towel	0.00E+00	0.00E+00	0.00E+00	8.09E-08	0.00E+00	8.86E-04	8.46E-06
137	tower	0.00E+00	0.00E+00	0.00E+00	3.84E-06	0.00E+00	8.84E-03	1.39E-04

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
138	trade name	0.00E+00	0.00E+00	0.00E+00	1.64E-05	0.00E+00	1.57E-02	3.40E-04
139	traffic light	0.00E+00	0.00E+00	0.00E+00	2.83E-05	0.00E+00	1.23E-02	3.80E-04
140	tray	0.00E+00	0.00E+00	0.00E+00	3.55E-08	0.00E+00	3.86E-04	3.69E-06
141	tree	2.23E-03	1.16E-01	2.17E-01	2.43E-01	3.48E-01	8.27E-01	1.54E-01
142	truck	0.00E+00	0.00E+00	0.00E+00	3.56E-04	0.00E+00	1.91E-01	4.02E-03
143	van	0.00E+00	0.00E+00	0.00E+00	1.64E-04	0.00E+00	6.99E-02	1.71E-03
144	vase	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
145	wall	0.00E+00	0.00E+00	1.20E-04	2.62E-03	1.52E-03	4.56E-01	1.07E-02
146	wardrobe	0.00E+00	0.00E+00	0.00E+00	3.09E-08	0.00E+00	3.39E-04	3.24E-06
147	washer	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
148	water	0.00E+00	0.00E+00	0.00E+00	4.62E-04	0.00E+00	9.80E-02	3.27E-03
149	waterfall	0.00E+00	0.00E+00	0.00E+00	9.88E-06	0.00E+00	1.64E-02	2.70E-04
150	windowpane	0.00E+00	0.00E+00	0.00E+00	2.12E-06	0.00E+00	5.39E-03	7.06E-05

Table S3. Model results for BC estimation using theory-driven feature selection

No.	All variables		Theory-driven: Outdoor variables		Theory-driven: Air pollution-related variables	
1	variable	coefficient	variable	coefficient	variable	coefficient
2	building_250	1.48E+00	building_250	7.28E-01	building_250	6.38E-01
3	pool table_1750	1.87E+06	sea_1250	3.32E+04	traffic light_1000	1.01E+03
4	stool_1500	-1.16E+07	ship_1000	4.49E+03	plant_500	-4.85E+00
5	ashcan_1250	1.65E+02	road_250	1.37E+00	bridge_250	5.02E+01
6	hovel_2000	-6.51E+03	swimming pool_1250	6.21E+03	road_central	4.18E-01
7	bag_250	-2.10E+03	sand_1750	-1.54E+03	truck_2000	-1.88E+02
8	sand_500	5.17E+02	tent_250	8.37E+01	mountain_750	-1.81E+02
9	ship_250	5.20E+02	fence_1250	-3.27E+01	sand_500	4.34E+02
10	tent_250	5.73E+01	fountain_750	-4.62E+03	land_1000	1.34E+04
11	rock_250	-1.18E+02	awning_500	2.22E+02	streetlight_250	1.91E+02
12	cabinet_250	-1.50E+03	minibike_central	1.15E+02		
13	canopy_1000	4.11E+03	lake_750	2.28E+03		
14	tower_250	1.29E+03	hill_central	7.42E+01		
15	painting_central	1.74E+04	tower_250	1.03E+03		
16	sconce_1000	-7.73E+04	skyscraper_500	-3.03E+05		
17	wall_250	1.46E+01	hovel_250	-1.25E+03		
18	shelf_1500	1.45E+05	river_750	-1.19E+02		
19	waterfall_1750	-5.18E+03	runway_500	-9.76E+02		
20	flower_750	-3.08E+02	mountain_250	-3.75E+01		
21	poster_750	8.71E+03				
Intercept	1.10E+00		8.49E-01		1.03E+00	
adjusted R^2	0.77		0.72		0.63	
10-fold CV R^2	0.42		0.61		0.57	
RMSE ($\mu g/m^3$)	0.20		0.22		0.26	

Table S4. Model results for PN estimation using theory-driven feature selection

No.	All variables		Theory-driven: Outdoor variables		Theory-driven: Air pollution-related variables	
1	variable	coefficient	variable	coefficient	variable	coefficient
2	river_2000	-2.29E+03	river_2000	-2.13E+03	streetlight_250	2.20E+02
3	signboard_250	6.22E+01	signboard_250	6.74E+01	truck_2000	9.36E+02
4	counter_1250	9.11E+03	grass_central	-8.95E-01	bridge_250	1.16E+02
5	awning_250	3.04E+02	ship_750	6.70E+03	traffic light_750	2.43E+03
6	bed_250	-5.99E+05	rock_2000	2.12E+03	grass_1000	4.84E+00
7	ship_750	5.27E+03	building_central	1.52E+00	path_central	-3.70E+00
8	grass_250	-2.11E+00	tent_250	2.42E+02	building_central	1.52E+00
9	food_750	-5.35E+03	sand_1500	-2.19E+03	sand_250	-5.50E+02
10	conveyer belt_1250	-5.83E+04	dirt track_1000	5.75E+02	plant_250	9.69E+00
11	building_central	1.19E+00	wall_250	2.23E+01	house_250	-3.31E+00
12	tent_250	1.93E+02	canopy_1000	5.84E+03	wall_250	2.38E+01
13	path_central	-3.41E+00	animal_250	9.06E+03	land_250	-3.93E+03
14	light_1500	9.72E+04			road_250	1.35E+00
15	desk_central	3.40E+04			car_250	4.73E+00
16	pot_central	2.58E+02			flower_1000	8.22E+02
17					mountain_250	6.32E+01
18					runway_1750	4.88E+03
Intercept	8.97E+00		8.44E+00		7.03E+00	
adjusted R^2	0.77		0.74		0.73	
10-fold CV R^2	0.57		0.65		0.65	
RMSE (pt/cm^3)	821		916		954	

Table S5. Number of input predictor variables offered to the buffer-feature models using different feature selection methods

Feature selection method	Variable types	Number of variables offered
None	All variables	1350
Theory-driven	Outdoor variables	513
Theory-driven	Air pollution related variables	279
Data-driven	All variables	252-659
Integrated	Outdoor variables	216-478
Integrated	Air pollution related variables	188-278

Table S6. Results of the best BC and PN buffer-feature models using data-driven feature selection

No.	BC		PN	
1	variable	coefficient	variable	coefficient
2	building_250	5.23E-01	river_2000	-2.50E+03
3	traffic light_1000	9.09E+02	signboard_250	5.49E+01
4	plant_500	-7.10E+00	counter_1250	1.19E+04
5	ashcan_1250	6.36E+02	awning_250	3.49E+02
6	road_central	1.94E-01	grandstand_1250	-3.98E+03
7	door_500	-9.14E+02	grass_250	-2.51E+00
8	water_1250	1.18E+02	field_500	1.81E+01
9	railing_1000	-5.17E+02	sand_250	-4.74E+02
10	signboard_500	1.39E+02	path_central	-3.08E+00
11	rock_250	-1.43E+02	house_central	-2.26E+00
12	streetlight_central	2.69E+01	canopy_1000	1.04E+04
13	grandstand_1750	3.49E+03	chair_2000	-2.08E+04
14	fence_750	-2.37E+01	building_central	8.54E-01
15	sand_500	3.39E+02	bathtub_1500	-2.69E+04
16	truck_1000	-1.10E+02	pot_central	2.40E+02
17	grass_250	-1.10E+00	door_500	-1.45E+03
18	hill_1500	-8.43E+01	wall_250	1.33E+01
19			tent_250	8.21E+01
20			van_central	-1.12E+01
21			land_1500	-8.59E+03
22			hovel_500	-9.56E+02
23			streetlight_central	4.52E+01
Intercept	1.12E+00		9.21E+00	
adjusted R^2	0.72		0.80	
10-fold CV R^2	0.60		0.70	
RMSE	0.22 ($\mu g/m^3$)		807 (pt/cm^3)	

Table S7. Summary of buffer-feature models and reduced buffer-feature models for air pollution estimation (all models used the integrated theory- and data-driven feature selection and the air pollution-related predictor variables)

Pollutants	Model type	Predictor variables	Number of selected variables	Adjusted R^2	10-fold CV R^2	RMSE
BC ($\mu g/m^3$)	Buffer-feature model	central (50m) + buffer (250-2000m)	9~12	0.57~0.64	0.50~0.57	0.25-0.28
	Reduced buffer-feature models	buffer (250-2000m)	9~15	0.59~0.65	0.48~0.59	0.25-0.27
		buffer (500-2000m)	11~12	0.59~0.60	0.51~0.53	0.27-0.27
		buffer (750-2000m)	7~12	0.51~0.58	0.48~0.51	0.27-0.30
		buffer (1000-2000m)	4~12	0.36~0.56	0.34~0.50	0.28-0.34
PN (pt/cm^3)	Buffer-feature model	central (50m) + buffer (250-2000m)	11~17	0.65~0.73	0.61~0.66	941-1030
	Reduced buffer-feature models	buffer (250-2000m)	11~20	0.66~0.72	0.60~0.64	980-1076
		buffer (500-2000m)	10~14	0.55~0.62	0.48~0.56	1127-1213
		buffer (750-2000m)	6~9	0.48~0.53	0.42~0.49	1264-1329
		buffer (1000-2000m)	4~6	0.35~0.47	0.32~0.43	1324-1475

Table S8. Results of the Global Moran's I test for 4 types of GSV-only buffer-feature models

Pollutants	Models	Inverse Distance				Inverse Distance Squared				Fixed Distance			
		300m		500m		300m		500m		300m		500m	
		Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value
BC ($\mu g/m^3$)	All variables	0.15	0.00	0.09	0.00	0.23	0.00	0.19	0.00	0.07	0.01	0.02	0.19
	Theory-driven: Air pollution related variables	0.32	0.00	0.22	0.00	0.40	0.00	0.34	0.00	0.25	0.00	0.12	0.00
	Data-driven: All variables	0.18	0.00	0.09	0.00	0.26	0.00	0.21	0.00	0.09	0.00	0.00	0.86
	Integrated: air pollution related variables	0.32	0.00	0.22	0.00	0.40	0.00	0.34	0.00	0.25	0.00	0.12	0.00
PN (pt/cm^3)	All variables	0.15	0.00	0.10	0.00	0.21	0.00	0.18	0.00	0.09	0.00	0.04	0.02
	Theory-driven: Air pollution related variables	0.21	0.00	0.15	0.00	0.25	0.00	0.22	0.00	0.16	0.00	0.09	0.00
	Data-driven: All variables	0.06	0.03	0.02	0.22	0.11	0.00	0.09	0.01	0.02	0.46	-0.03	0.20
	Integrated: air pollution related variables	0.22	0.00	0.16	0.00	0.26	0.00	0.22	0.00	0.17	0.00	0.10	0.00

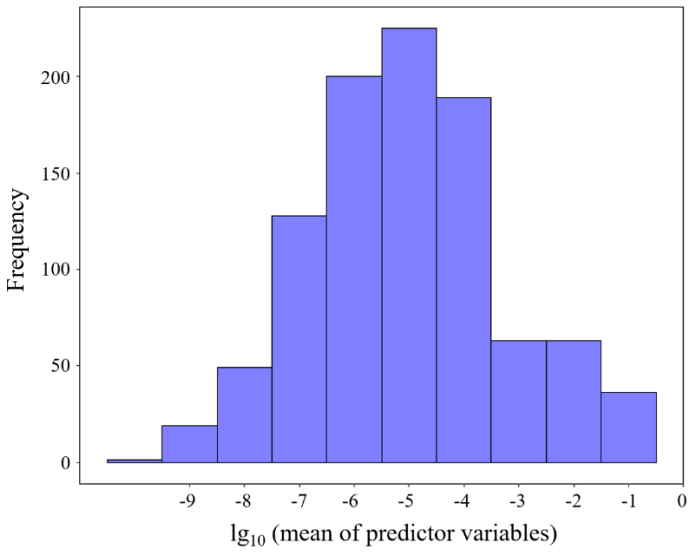

Figure S1. Histogram of the mean of 1,350 predictor variables. The values were log-transformed to determine the magnitude of thresholds for data filtering.

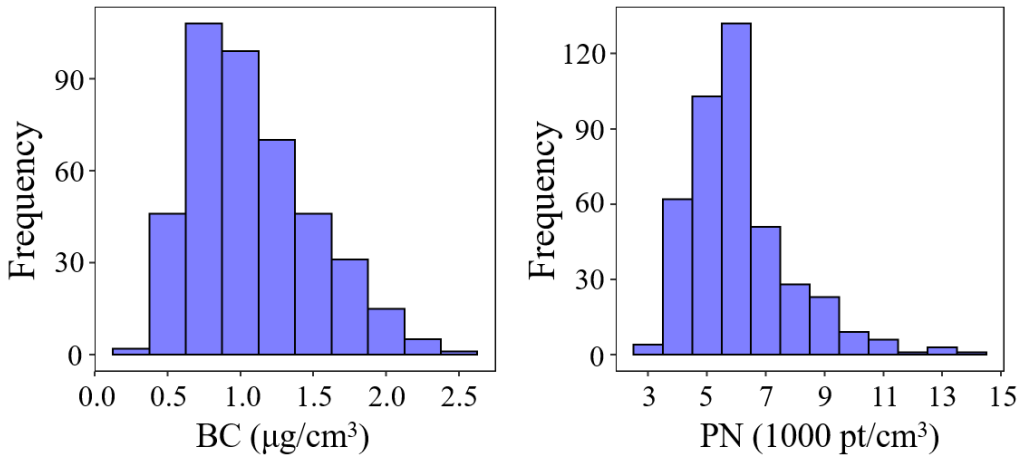

Figure S2. Histograms of the multiplicative background adjusted BC and PN concentrations.

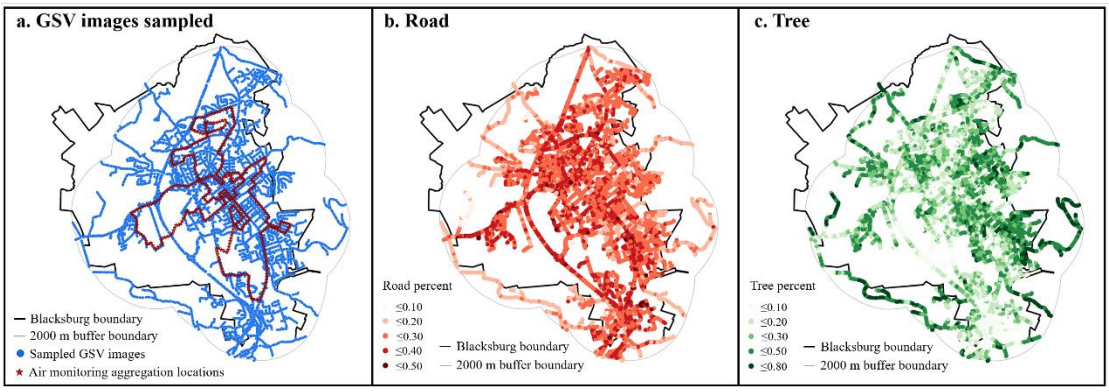

Figure S3. Spatial distribution of the sampled GSV images and examples of built and natural environment characteristics derived from the images. (a) GSV images sampled, (b) road distribution, and (c) tree cover distribution

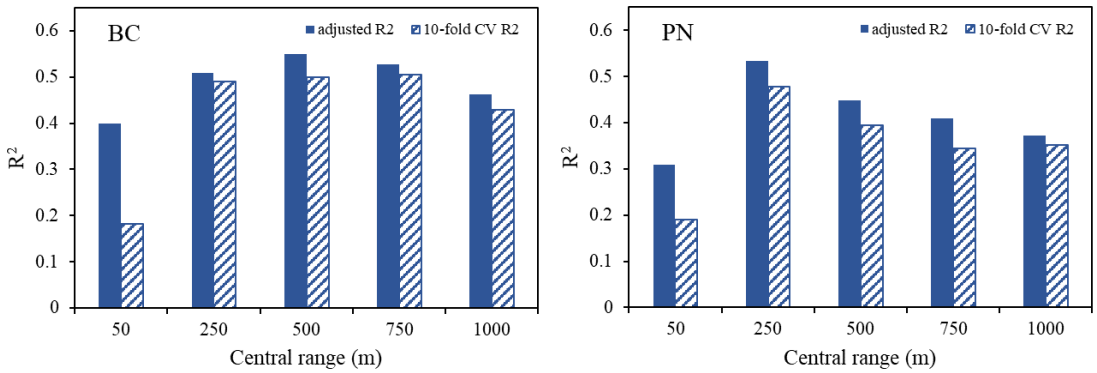

Figure S4. Comparison of central-feature models with different distance ranges. The predictor variables for both models were "Theory-driven: Air pollution-related variables".

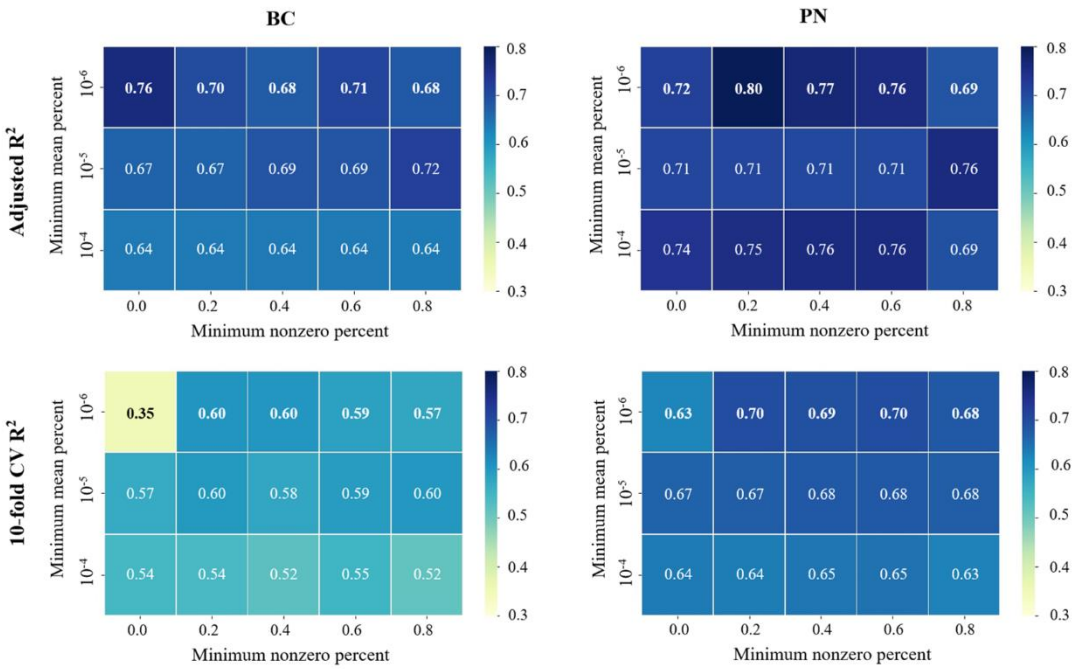

Figure S5. Model performance of buffer-feature models among different scenarios using data-driven feature selection. The upper panel shows adjusted R^2 of the models and the bottom panel shows 10-fold cross validation R^2. Analogous plots for the central-feature models are in Figure S6.

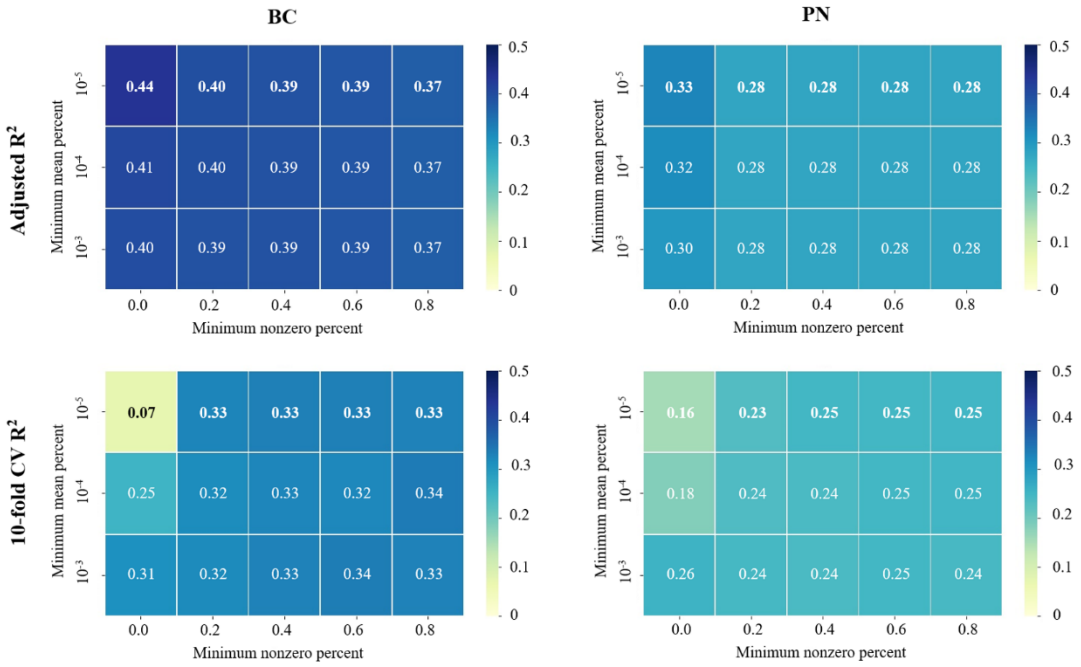

Figure S6. Model performance of the central-feature models among different scenarios using the data-driven feature selection. The upper panel shows adjusted R^2 of the models and the bottom panel shows 10-fold cross validation R^2.

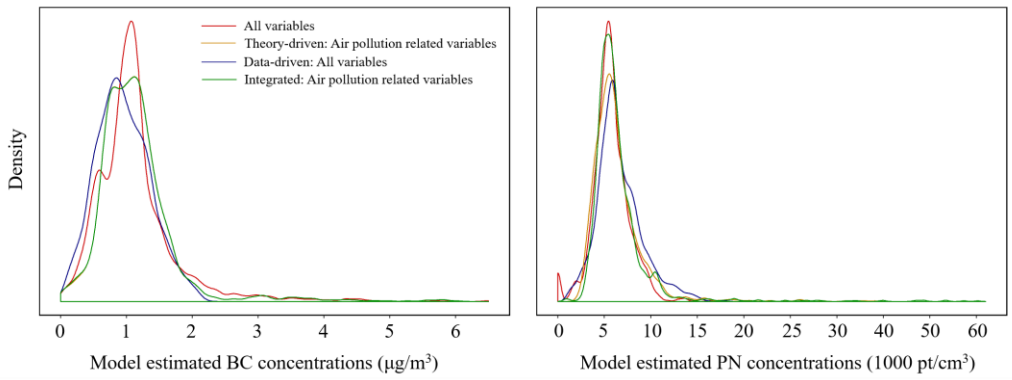

Figure S7. Density plots of estimated BC and PN concentrations using 4 types of GSV-only models. All distributions were similar and highly positive-skewed.

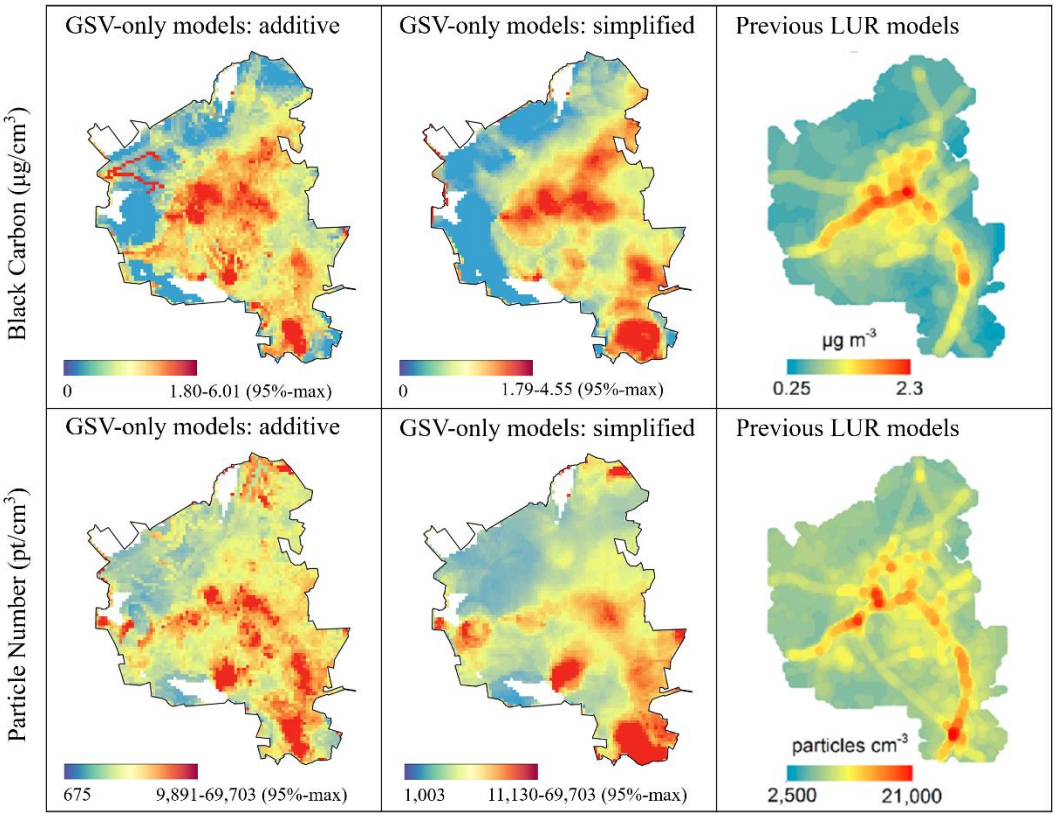

Figure S8. Comparison among GSV-only model estimates and previous LUR model estimates. Figures from the left to right columns were estimates using reduced buffer-feature models with the additive method, estimates using reduced models with the simplified method, and estimates using previous LUR models with traditional predictor variables[1]. Due to highly right-skewed distributions of the predicted concentrations, the color bar was scaled from minimum to the 95th percentile to better visualize spatial patterns of the data.

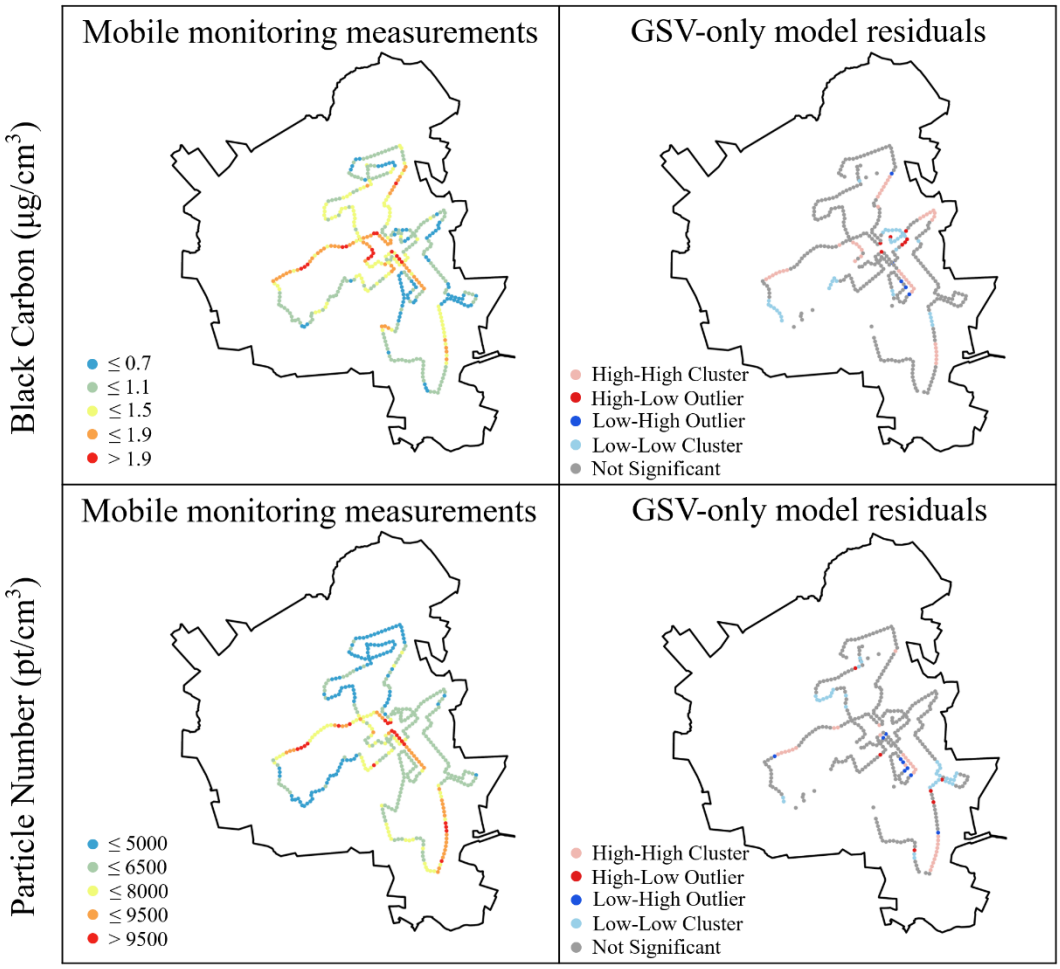

Figure S9. The spatial distribution of mobile monitoring measurements and the results of LISA test for GSV-only model residuals. We used the integrated buffer-feature models with air pollution-related variables to illustrate the results of the LISA analysis (500m threshold). High-High and Low-Low cluster represent statistically significant clusters of underestimates and overestimates of particulate air pollution – these clusters are the most important for the models presented in this paper. A High-Low (Low-High) outlier are where a high (low) residual value is surrounded primarily by low (high) residual values.

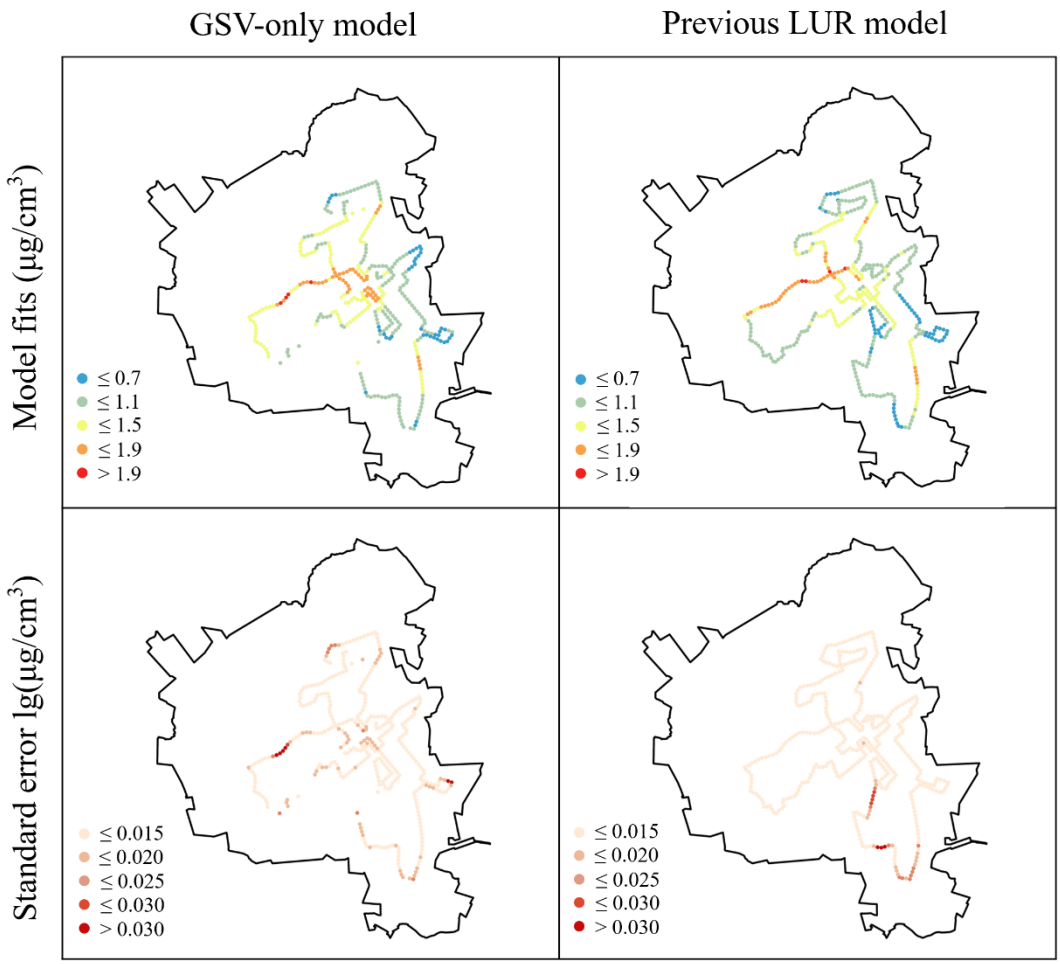

Figure S10. Comparison of BC model uncertainty for the GSV-only model presented in this paper and the previously published LUR model with traditional GIS predictor variables[1]. We used the integrated buffer-feature model with air pollution-related variables as an example. The upper panel shows model fits; the bottom panel shows standard error for model prediction. The model fits are transformed back to original values for illustration.

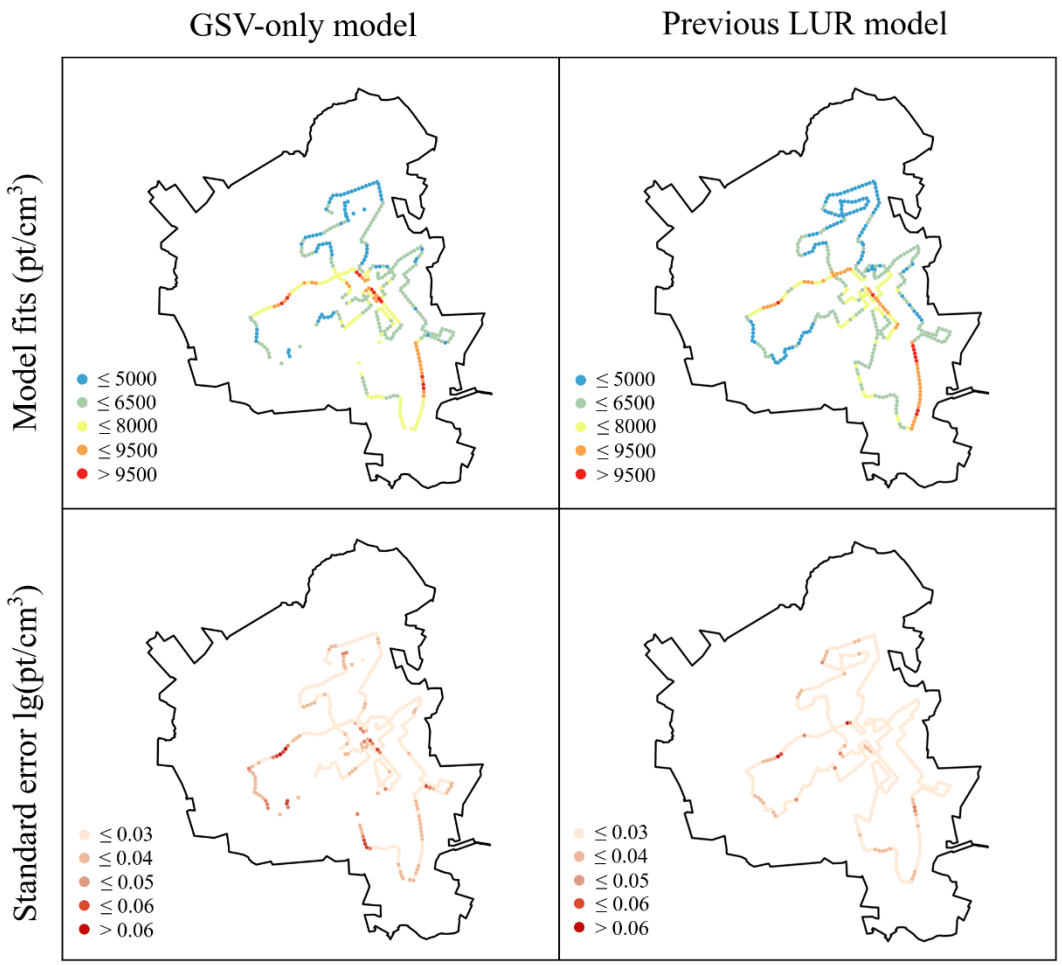

Figure S11. Comparison of PN model uncertainty for the GSV-only model presented in this paper and the previously published LUR model with traditional GIS predictor variables[1]. We used the integrated buffer-feature model with air pollution-related variables as an example. The upper panel shows model fits; the bottom panel shows standard error for model prediction. The model fits are transformed back to original values for illustration.

REFERENCE

1. Hankey, S.; Sforza, P.; Pierson, M., Using Mobile Monitoring to Develop Hourly Empirical Models of Particulate Air Pollution in a Rural Appalachian Community. *Environ. Sci. Technol.* **2019,** *53*, 4305-4315.